



**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Patent Application of

Ogasawara et al.

Serial No.: 10/787,005

Filed: February 25, 2004

For: COMPILER DEVICE, PROGRAM, AND RECORDING MEDIUM

Date: December 7, 2004

Group Art Unit: 2655

Examiner: Not yet assigned

Docket No.: JP920030021US1

Mail Stop Amendment  
Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

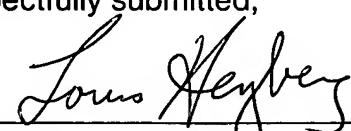
**SUBMISSION OF PRIORITY DOCUMENT**

Sir:

Enclosed herewith is a certified copy of Japanese Application No. 2003-49414 filed February 26, 2003, in support of applicant's claim to priority under 35 U.S.C. 119.

Respectfully submitted,

By



Louis P. Herzberg

Reg. No. 41,500

Phone No. (914) 945-2885

IBM Corporation  
Intellectual Property Law Dept.  
P.O. Box 218  
Yorktown Heights, NY 10598

日本国特許庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されて  
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed  
with this Office

出願年月日

Date of Application: 2003年 2月 26日

出願番号

Application Number: 特願 2003-049414

ST.10/C]:

[JP 2003-049414]

出願人

Applicant(s): インターナショナル・ビジネス・マシンズ・コーポレーション

2003年 5月 20日

特許庁長官  
Commissioner,  
Japan Patent Office

太田 信一郎

出証番号 出証特 2003-3037028

【書類名】 特許願

【整理番号】 JP9030021

【提出日】 平成15年 2月26日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 9/45

【発明者】

【住所又は居所】 神奈川県大和市下鶴間1623番地14 日本アイ・ビー・エム株式会社 東京基礎研究所内

【氏名】 竹内 幹雄

【発明者】

【住所又は居所】 神奈川県大和市下鶴間1623番地14 日本アイ・ビー・エム株式会社 東京基礎研究所内

【氏名】 小野寺 民也

【発明者】

【住所又は居所】 神奈川県大和市下鶴間1623番地14 日本アイ・ビー・エム株式会社 東京基礎研究所内

【氏名】 小笠原 武史

【特許出願人】

【識別番号】 390009531

【氏名又は名称】 インターナショナル・ビジネス・マシンズ・コーポレーション

【代理人】

【識別番号】 100086243

【弁理士】

【氏名又は名称】 坂口 博

【代理人】

【識別番号】 100091568

【弁理士】

【氏名又は名称】 市位 嘉宏

【代理人】

【識別番号】 100108501

【弁理士】

【氏名又は名称】 上野 剛史

【複代理人】

【識別番号】 100104156

【弁理士】

【氏名又は名称】 龍華 明裕

【手数料の表示】

【予納台帳番号】 053394

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9706050

【包括委任状番号】 9704733

【包括委任状番号】 0207860

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 コンパイラ装置、コンパイラプログラム、及び記録媒体

【特許請求の範囲】

【請求項1】 文字列を操作するプログラムを最適化するコンパイラ装置であって、

前記プログラムにおいて、文字列を格納する文字列変数に文字列を追加する追加命令を検出する追加命令検出部と、

前記追加命令検出部により検出された、同一の文字列変数に文字列を追加する複数の追加命令のそれぞれに代えて、当該追加命令により追加される追加文字列のデータをバッファに格納する格納コードを生成する格納コード生成部と、

前記プログラムにおける、前記文字列変数を参照する命令より先に実行される個所に、複数の前記追加文字列のそれぞれを前記文字列変数に追加する追加コードを生成する追加コード生成部と

を備えるコンパイラ装置。

【請求項2】 前記複数の追加命令によって文字列を追加した後に初めて前記文字列変数を参照する参照命令を検出する参照命令検出部を更に備え、

前記追加コード生成部は、前記追加コードを、全ての前記格納コードより後、かつ前記参照命令より先に実行される個所に生成する請求項1記載のコンパイラ装置。

【請求項3】 前記追加命令検出部は、前記追加命令として、文字列を追加する処理が許可されていない不变文字列変数を、文字列を追加する処理が許可された可変文字列変数に変換する命令と、前記可変文字列変数に前記追加文字列を追加する命令と、前記可変文字列変数を不变文字列変数に変換する命令との組合せを検出する請求項1記載のコンパイラ装置。

【請求項4】 文字列を操作するプログラムを最適化するコンパイラ装置であって、

前記プログラムにおいて、文字列を格納する文字列変数に文字列を追加する追加命令を検出する追加命令検出部と、

前記追加命令検出部により検出された、同一の文字列変数に文字列を追加する

複数の追加命令のそれぞれに代えて、当該追加命令により追加される追加文字列が格納されているメモリのアドレスをバッファに格納する格納コードを生成する格納コード生成部と、

前記プログラムにおける、前記文字列変数を参照する命令より先に実行される個所に、複数の前記アドレスに格納される複数の追加文字列のそれぞれを前記文字列変数に追加する追加コードを生成する追加コード生成部とを備えるコンパイラ装置。

【請求項5】 文字列を操作するプログラムを最適化するコンパイラ装置であって、

文字列を追加する処理が許可された可変文字列変数を、文字列を追加する処理が許可されていない不变文字列変数に変換する不変化命令を検出する不変化命令検出部と、

前記不变文字列変数を可変文字列変数に変換する可変化命令を検出する可変化命令検出部と、

前記不変化命令から前記可変化命令までの間に実行される命令が、前記不変化命令の変換元の可変文字列変数に格納された文字列を書き換えず、かつ前記可変化命令から前記可変化命令により変換された可変文字列変数の使用の間に実行される命令が、前記不変化命令の変換元の可変文字列変数と、前記可変化命令により変換された可変文字列変数の何れも書き換えない場合に、前記可変化命令を除去し、前記可変化命令より後で、前記不変化命令の変換元となる可変文字列変数を、前記可変化命令により変換された可変文字列変数として使用させる命令除去部と

を備えるコンパイラ装置。

【請求項6】 前記不变文字列変数に格納された文字列が参照されない場合に、前記命令除去部は、前記不変化命令を更に除去する請求項5記載のコンパイラ装置。

【請求項7】 前記命令除去部は、前記不変化命令を、当該不変化命令より後に実行される分岐命令の分岐先のそれぞれに移動し、当該分岐命令の分岐先のそれぞれにおいて、不変化命令の変換先である不变文字列変数に格納されている

文字列が参照されない場合に、当該不変化命令を除去する部分不用代入文除去を行う請求項6記載のコンパイラ装置。

【請求項8】 前記可変化命令検出部は、可変文字列変数として用いられる記憶領域を確保する命令と、当該可変文字列変数に、前記不変文字列変数に格納された文字列を追加する命令との組合せを、前記可変化命令として検出する請求項5記載のコンパイラ装置。

【請求項9】 前記可変化命令検出部により検出された前記可変化命令を、当該可変化命令より前において合流する制御フローの合流元のそれぞれに移動する部分冗長性の除去処理を行う部分冗長除去部を更に備え、

前記部分冗長性の前記除去処理が実行されたプログラムにおいて、前記不変化命令及び前記可変化命令の間に実行される命令が、前記不変化命令の変換元の可変文字列変数に格納された文字列を書き換えず、かつ前記可変化命令から前記可変化命令により変換された可変文字列変数の使用の間に実行される命令が、前記不変化命令の変換元の可変文字列変数と、前記可変化命令により変換された可変文字列変数の何れも書き換えない場合に、前記命令除去部は、前記可変化命令を除去する請求項5記載のコンパイラ装置。

【請求項10】 前記命令除去部は、前記不変化命令を、当該不変化命令より後に実行される分岐命令の分岐先のそれぞれに移動し、当該分岐命令の分岐先のそれぞれにおいて、不変化命令の変換先である不変文字列変数に格納されている文字列が参照されない場合に、当該不変化命令を除去する部分不用代入文除去を行う請求項9記載のコンパイラ装置。

【請求項11】 文字列を操作するプログラムをコンピュータにより最適化するコンパイラプログラムであって、

前記コンピュータを、

前記プログラムにおいて、文字列を格納する文字列変数に文字列を追加する追加命令を検出する追加命令検出部と、

前記追加命令検出部により検出された、同一の文字列変数に文字列を追加する複数の追加命令のそれぞれに代えて、当該追加命令により追加される追加文字列のデータをバッファに格納する格納コードを生成する格納コード生成部と、

前記プログラムにおける、前記文字列変数を参照する命令より先に実行される個所に、複数の前記追加文字列のそれぞれを前記文字列変数に追加する追加コードを生成する追加コード生成部と  
して機能させるコンパイラプログラム。

【請求項12】 文字列を操作するプログラムをコンピュータにより最適化するコンパイラプログラムであって、

前記コンピュータを、

前記プログラムにおいて、文字列を格納する文字列変数に文字列を追加する追加命令を検出する追加命令検出部と、

前記追加命令検出部により検出された、同一の文字列変数に文字列を追加する複数の追加命令のそれぞれに代えて、当該追加命令により追加される追加文字列が格納されているメモリのアドレスをバッファに格納する格納コードを生成する格納コード生成部と、

前記プログラムにおける、前記文字列変数を参照する命令より先に実行される個所に、複数の前記アドレスに格納される複数の追加文字列のそれぞれを前記文字列変数に追加する追加コードを生成する追加コード生成部と  
して機能させるコンパイラプログラム。

【請求項13】 文字列を操作するプログラムをコンピュータにより最適化するコンパイラプログラムであって、

前記コンピュータを、

文字列を追加する処理が許可された可変文字列変数を、文字列を追加する処理が許可されていない不变文字列変数に変換する不变化命令を検出する不变化命令検出部と、

前記不变文字列変数を可変文字列変数に変換する可変化命令を検出する可変化命令検出部と、

前記不变化命令から前記可変化命令までの間に実行される命令が、前記不变化命令の変換元の可変文字列変数に格納された文字列を書き換えず、かつ前記可変化命令から前記可変化命令により変換された可変文字列変数の使用の間に実行される命令が、前記不变化命令の変換元の可変文字列変数と、前記可変化命令によ

り変換された可変文字列変数の何れも書き換えない場合に、前記可変化命令を除去し、前記可変化命令より後で、前記不変化命令の変換元となる可変文字列変数を、前記可変化命令により変換された可変文字列変数として使用させる命令除去部と

して機能させるコンパイラプログラム。

【請求項14】 請求項11から13の何れかに記載のコンパイラプログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、コンパイラ装置、コンパイラプログラム、及び記録媒体に関する。特に本発明は、文字列を操作するプログラムを最適化するコンパイラ装置、コンパイラプログラム、及び記録媒体に関する。

【0002】

【従来の技術】

近年、プログラムの保守性や堅牢性を高めるために、プログラマにとって直感的な演算子を記載することのできるプログラム言語が普及している。例えば、Java（登録商標）言語において、プログラマは、文字列変数に文字列を追加する処理を、「+」演算子により、簡易に記載することができる。

また、従来、コンパイラにおいて、部分冗長性を除去する技術（非特許文献1参照。）及び部分不用代入文を除去する技術（非特許文献2参照。）が用いられている。

【0003】

【非特許文献1】

J. Knoopp, O. Ruthing and B. Steffen  
、Lazy code motion, In PLDI '92, p. 224-234, 1992. 邦題「Lazyコード移動」

【0004】

【非特許文献2】

J. Knoop, O. Ruthing and B. Steffen  
、Partial dead code elimination, In PL  
DI '94, p. 147-158, 1994. 邦題「部分不用代入文除去」

【0005】

【発明が解決しようとする課題】

文字列の追加処理は、実際には、メモリ領域の確保及びデータの複写等の、多数の処理の集まりにコンパイルされる。しかしながら、同一の文字列変数に、複数の文字列が順次追加される場合には、メモリ領域の確保及びデータの複写等が冗長となる場合がある。更に、これらの冗長な命令は、部分冗長性の除去及び部分不用代入文の除去の技術をそのまま適用しても、適切に除去されない場合がある。例えば、サーバ用プログラムにおいて、文字列の追加処理が頻繁に用いられる場合には、サーバの動作効率を低めてしまっていた。

そこで本発明は、上記の課題を解決することのできるコンパイラ装置、コンパイラプログラム、及び記録媒体を提供することを目的とする。この目的は特許請求の範囲における独立項に記載の特徴の組み合わせにより達成される。また従属項は本発明の更なる有利な具体例を規定する。

【0006】

【課題を解決するための手段】

即ち、本発明の第1の形態によると、文字列を操作するプログラムを最適化するコンパイラ装置であって、プログラムにおいて、文字列を格納する文字列変数に文字列を追加する追加命令を検出する追加命令検出部と、追加命令検出部により検出された、同一の文字列変数に文字列を追加する複数の追加命令のそれぞれに代えて、当該追加命令により追加される追加文字列のデータをバッファに格納する格納コードを生成する格納コード生成部と、プログラムにおける、文字列変数を参照する命令より先に実行される個所に、複数の追加文字列のそれぞれを文字列変数に追加する追加コードを生成する追加コード生成部とを備えるコンパイラ装置、コンパイラプログラム、及びコンパイラプログラムを記録した記録媒体を提供する。

なお上記の発明の概要は、本発明の必要な特徴の全てを列挙したものではなく

、これらの特徴群のサブコンビネーションも又発明となりうる。

【0007】

【発明の実施の形態】

以下、発明の実施の形態を通じて本発明を説明するが、以下の実施形態は特許請求の範囲にかかる発明を限定するものではなく、又実施形態の中で説明されている特徴の組み合わせの全てが発明の解決手段に必須であるとは限らない。

【0008】

(第1実施形態)

図1は、コンパイラ装置10のブロック図を示す。コンパイラ装置10は、例えば、Java(登録商標)のJust-In-Timeコンパイラであり、文字列を格納する文字列変数に文字列を追加する追加命令を最適化することを目的とする。コンパイラ装置10は、追加命令を検出する追加命令検出部100と、追加命令により追加される追加文字列のデータをバッファに格納する格納コードを生成する格納コード生成部110と、文字列変数を参照する参照命令を検出する参照命令検出部120と、バッファに格納された追加文字列を文字列変数に追加する追加コード生成部130とを備える。

【0009】

追加命令検出部100は、コンパイル対象のプログラムを取得すると、当該プログラムにおいて、同一の文字列変数に文字列を追加する複数の追加命令を検出し、検出結果を格納コード生成部110及び参照命令検出部120に送る。格納コード生成部110は、追加命令の検出結果に基づき、コンパイル対象のプログラムにおける、追加命令検出部100によって検出された複数の追加命令のそれぞれに代えて、追加命令により追加される追加文字列のデータをバッファに格納する格納コードを生成する。そして、格納コード生成部110は、格納コードが生成された当該プログラムを追加コード生成部130に送る。

【0010】

参照命令検出部120は、追加命令の検出結果に基づき、コンパイル対象のプログラムにおいて、複数の追加命令によって文字列を追加した後初めて文字列変数を参照する参照命令を検出し、検出結果を追加コード生成部130に送る。追

加コード生成部130は、参照命令の検出結果に基づき、格納コード生成部110から受け取ったプログラムにおける、全ての格納コードより後、かつ参照命令より先に実行される個所に、複数の追加文字列のそれぞれを文字列変数に追加する追加コードを生成する。そして、追加コード生成部130は、追加コードが生成された当該プログラムを、コンパイル結果のプログラムとして出力する。なお、コンパイラ装置10は、本図の各部材が行う処理より先又は後に、他の最適化を行ってもよい。

#### 【0011】

図2は、コンパイラ装置10の動作フローを示す。コンパイラ装置10は、例えば、Java（登録商標）における、コンパイル対象のプログラムを含むクラスファイルを取得する（S200）。そして、追加命令検出部100は、同一の文字列変数に文字列を追加する複数の追加命令を検出する（S210）。格納コード生成部110は、コンパイル対象のプログラムにおいて、追加命令検出部100において検出された複数の追加命令のそれぞれに代えて、追加命令により追加される追加文字列のデータをバッファに格納する格納コードを生成する（S220）。参照命令検出部120は、複数の追加命令によって文字列を追加した後に初めて文字列変数を参照する参照命令を検出する（S230）。そして、追加コード生成部130は、全ての格納コードより後、かつ参照命令より先に実行される個所に、複数の追加文字列のそれぞれを文字列変数に追加する追加コードを生成する（S240）。

#### 【0012】

図3（a）は、最適化対象のプログラムのソースコードの一例を示す。1行目の命令は、文字列変数の一例であるsに、空の文字列“”を代入する命令である。2から4行目の命令のそれぞれは、変数cに格納された追加文字列をsと連結(concatenate)し、その結果をsに代入する命令、即ち、変数cに格納された追加文字列をsに追加する追加命令である。なお、2から4行目の命令のそれぞれは、ループにより繰り返し実行される、ソースコード上で同一の追加命令であってもよい。5行目の命令は、2から4行目のそれぞれの追加命令によって文字列を追加した後に初めてsを参照する参照命令である。

## 【0013】

図3 (b) は、プログラムがコンパイルされた結果を示す。より詳細には、例えば、本図は、Java (登録商標) におけるバイトコードコンパイラが、図3 (a) に示したソースコードを、バイトコードにコンパイルした結果の一例を示す。なお、実際には、バイトコードコンパイラは、各命令を数値データで示したバイトコードを出力するが、本図は、説明の都合上バイトコードの意味を示す擬似コードを示す。なお、以降に説明する図5及び図8から12においても、説明の都合上、命令の意味を示す擬似コードを用いる。

## 【0014】

1行目の命令は、文字列の追加が許可されていない不变文字列変数 `s` を、文字列の追加が許可された可変文字列変数 `s b` に変換する命令であり、例えば、文字列の追加が許可された可変文字列変数 `s b` を格納する領域をメモリに確保し、文字列の追加が許可されていない不变文字列変数 `s` に格納される文字列を、`s b` に複写する。2行目の命令は、`s b` に、変数 `c` に格納された追加文字列を追加する命令である。なお、変数 `c` がキャラクタ変数でない場合には、2行目の命令に加え、更に、`valueOf` メソッドが実行され、その過程で不变文字列変数が更に確保される場合がある。3行目の命令は、`s b` を、不变文字列 `s` に変換する命令である。以降同様に、4行目及び7行目の命令のそれぞれは、1行目の命令と略同一であり、5行目及び8行目の命令のそれぞれは、2行目の命令と略同一であり、6行目及び9行目の命令のそれぞれは、3行目の命令と略同一である。また、10行目の命令は、不变文字列変数 `s` を参照する命令である。

## 【0015】

このように、不变文字列変数 `s` に、新たに文字列 `c` を追加する命令は、メモリ上の領域確保及びデータの複写を含む複数の命令にコンパイルされる。本図のプログラムによると、最終的に用いられない変数の領域確保が、ループで繰り返されてしまう。例えば、1行目、4行目、及び7行目において、最終的に不要な変数 `s b` の領域確保が繰り返される。更に、新たに確保されたメモリ領域を初期化する処理が必要となり、効率が悪い。また、コンパイラ装置10が動的メモリ管理を行う場合には、不要なメモリ領域は、ガーベージコレクション (Garbage Collection) によって解放される。

ge Collection) により頻繁に開放され、効率が悪い。

#### 【0016】

ここで、文字列を追加する処理が許可された可変文字列変数 (mutable string variable) とは、プログラム言語の仕様により、文字列の追加操作が許可されている変数であり、一例としては、Java (登録商標) 言語における StringBuffer オブジェクトのインスタンスである。そして、プログラム言語がオブジェクト指向言語の場合、可変文字列変数を含む可変文字列オブジェクトは、可変文字列変数に格納される文字列に新たに文字列を追加する追加メソッド、例えば、append メソッドを有していてもよい。

#### 【0017】

一方、文字列を追加する処理が許可されていない不变文字列変数 (immutable string variable) とは、プログラム言語の仕様により、文字列の追加操作が許可されていない変数であり、一例としては、Java (登録商標) 言語における String オブジェクトのインスタンスである。そして、コンパイラ装置 10 は、不变文字列変数に格納された文字列が変化しないと仮定してコンパイル処理を行えるので、効率的となる。例えば、コンパイラ装置 10 は、不变文字列変数へのアクセスに、排他制御を必要としないので、高速に実行させることができる。

#### 【0018】

図3 (c) は、コンパイラ装置 10 による最適化を終えたプログラムの一例を示す。追加命令検出部 100 は、追加命令として、不变文字列変数 s を可変文字列変数 s\_b に変換する図3 (b) の1行目の命令と、可変文字列変数 s\_b に追加文字列 c を追加する図3 (b) の2行目の命令と、可変文字列変数 s\_b を不变文字列変数 s に変換する図3 (b) の3行目の命令との組合せを検出する。そして、格納コード生成部 110 は、追加命令に代えて、追加文字列 c のデータをバッファに格納する格納コード、例えば、図3 (c) の2行目の命令を生成する。

#### 【0019】

ここで、バッファとは、プログラムを実行するスレッド毎に、予め確保されたメモリ上の領域である。バッファは、本図のプログラムの実行に加え、同一のス

レッドによる他の処理に用いられるので、文字のデータサイズより充分に大きいことが望ましい。一例としては、バッファのサイズは、1 M b y t e であってもよいし、10 M b y t e であってもよい。また、本図で示した変数 b u f は、バッファにおける空き領域の先頭のアドレスを格納する変数である。

#### 【0020】

追加コード生成部130は、不变文字列変数 s を参照する7行目の命令より先に実行される箇所、例えば、5行目に、複数の追加文字列のそれぞれを文字列変数に追加する追加コードを生成する。例えば、追加コード生成部130は、追加コードとして、追加文字列を格納しているバッファのアドレスを、不变文字列変数を含むオブジェクトのインスタンス変数 (s. v a l u e) に代入する命令を生成する。この結果、7行目の命令は、追加処理を終えた後の文字列変数を適切に参照できる。更に、本図に示したプログラムを実行するスレッドが、本図に示したプログラム以降に更に他のプログラムを実行する場合に備え、追加コード生成部130は、バッファの空き領域を示す情報を更新する処理を行うコードを生成することが望ましい。例えば、追加コード生成部130は、変数 b u f の値を更新する命令を6行目に生成することにより、後に実行されるコードに、バッファの空き領域を適切に知らせることができる。

#### 【0021】

このように、コンパイラ装置10は、可変文字列変数の領域をメモリに確保する命令及び文字列のデータを複写する命令を除去することにより、文字列を追加する追加命令を最適化することができる。

なお、本実施形態において、コンパイラ装置10は、例えば、バイトコードを実行コードにコンパイルする J u s t - I n - T i m e コンパイラである。これに代えて、コンパイラ装置10は、ソースプログラムをバイトコードにコンパイルするバイトコードコンパイラであってもよい。この場合、追加命令検出部100は、ソースプログラムにおいて、文字列変数に文字列を追加する「+」演算子を、追加命令として検出する。そして、格納コード生成部110は、「+」演算子に代えて、格納コードを生成する。即ち、コンパイラ装置10は、図3 (a) に示したソースプログラムを入力とし、図3 (b) のプログラムを介さずに、図

3 (c) と略同一の処理を行うバイトコードを出力してもよい。

#### 【0022】

図4は、変形例において、コンパイラ装置10がプログラムを最適化する一例を示す。本例におけるコンパイラ装置10は、図3 (c) に示したプログラムに代えて、本図に示したプログラムを生成してもよい。

#### 【0023】

コンパイラ装置10は、バッファを初期化する1行目の命令を生成する。バッファbufは、連結リスト構造であり、連結リストの各要素は、追加文字列を1づつ格納することができる。格納コード生成部110は、図3 (a) における追加命令に代えて、追加文字列cのデータをバッファに格納する図4の2行目の命令と、連結リストにおける次の要素を示すアドレスを取得する3行目の命令と、格納コードとして生成する。以降同様に、格納コード生成部110は、複数の追加命令のそれぞれに代えて、4行目及び6行目において、格納コードを生成する。そして、追加コード生成部130は、不变文字列変数sを参照する9行目の命令より先に実行される個所、例えば、8行目に、複数の追加文字列のそれぞれを文字列変数に追加する追加コードを生成する。

#### 【0024】

このように、格納コード生成部110は、複数の追加命令のそれぞれに代えて、追加文字列及び追加文字列が格納されているアドレスとの組合せを、連結リスト状に、バッファに格納してもよい。この場合、連結リストをヌルポインタで予め終端しておくことで、コンパイラ装置10は、メモリアクセス違反を検出することにより、連結リストの終端を検出できる。これにより、各格納コードにおける、連結リストの終端チェック命令が省略できるので、処理を高速化することができる。

#### 【0025】

図5 (a) は、第1の他の方法においてプログラムが最適化された結果を示す。より詳細には、本例は、熟練したJava (登録商標) プログラマが、文字列操作の実行効率を高めるべく、「+」演算子を用いることなく、文字列の追加処理を記述した例である。1行目の命令は、可変文字列変数s\_bを格納する領域を

メモリに確保し、不变文字列変数 s に格納される文字列を、 s b に複写する命令である。2行目の命令は、 s b に、変数 c に格納された追加文字列を追加する命令である。以降同様に、3行目及び4行目の命令のそれぞれは、2行目の命令と略同一である。そして、5行目の命令は、 s b を、不变文字列 s に変換する命令である。また、6行目の命令は、不变文字列変数 s を参照する命令である。

#### 【0026】

このように、本例によると、追加命令に代えて、可変文字列変数に文字列を追加するコードを用いるので、図3 (b) に示したプログラムより性能が高い。

#### 【0027】

図5 (b) は、第2の他の方法においてプログラムが最適化された結果を示す。本例において、不变文字列変数を含む String オブジェクトは、 Java (登録商標) の言語仕様に規定されていない、非公開文字列構築子を有している。非公開文字列構築子によると、新たな String オブジェクトを生成すると共に、新たな当該 String オブジェクトに、複数の String オブジェクトを連結した文字列を複写することができる。例えば、2から4行目の命令のそれぞれは、新たに不变文字列変数を生成し、当該不变文字列変数に、2つの不变文字列変数の内容を複写する命令である。

#### 【0028】

これにより、可変文字列変数への変換処理を省くことができるので、本例に示したプログラムは、図3 (b) に示したプログラムより性能が高い。しかしながら、2から4行目のそれぞれにおいて、不变文字列変数を格納する領域が確保されるにも関わらず、5行目の命令においては、4行目の命令により生成された不变文字列変数のみを参照するので、依然として冗長である。

#### 【0029】

これに対し、第1実施形態におけるコンパイラ装置10は、冗長なメモリ確保命令や変換処理を除去することにより、文字列を操作する命令を適切に最適化することができる。

#### 【0030】

(第2実施形態)

図6は、コンパイラ装置60のブロック図を示す。コンパイラ装置60は、例えば、Java（登録商標）のJust-In-Timeコンパイラであり、文字列を格納する文字列変数に文字列を追加する追加命令を最適化することを目的とする。コンパイラ装置60は、不变文字列変数を可変文字列変数に変換する可変化命令を検出する可変化命令検出部の一例である意味復元部600と、部分冗長性の除去処理を実行する部分冗長除去部610と、可変文字列変数を不变文字列変数に変換する不変化命令を検出する不変化命令検出部620と、可変化命令を除去する命令除去部630と、検出した可変化命令を実行可能な状態に変換する逆意味復元部640と、コンパイル対象のプログラムを実行することによりプロファイルデータを取得する計測コード付きプログラム実行部642と、プロファイルデータを格納する不变文字列変数最終長データベース645と、可変文字列変数の初期長を指定する可変文字列変数初期長指定部650とを備える。

#### 【0031】

意味復元部600は、コンパイル対象のプログラムを取得すると、不变文字列変数を可変文字列変数に変換する可変化命令を検出する。より詳細には、例えば、意味復元部600は、可変化命令の機能を実現する複数の命令の組合せを検出し、当該命令の組合せを、可変化命令を示す擬似命令に変換する。そして、意味復元部600は、変換した擬似命令を含むプログラムを、部分冗長除去部610に送る。

#### 【0032】

部分冗長除去部610は、擬似命令を含むプログラムを意味復元部600から受け取ると、意味復元部600により検出された可変化命令を、当該可変化命令より前において合流する制御フローの合流元のそれぞれに移動する部分冗長性の除去処理を行う。なお、移動とは、受け取ったプログラムにおける可変化命令を削除し、当該可変化命令を制御フローの合流元に挿入することにより、命令の実行順序を変更することである。そして、部分冗長除去部610は、部分冗長性の除去処理を行ったプログラムを不変化命令検出部620に送る。

#### 【0033】

不変化命令検出部620は、部分冗長除去部610から受け取ったプログラム

において、可変文字列変数を不变文字列変数に変換する不変化命令を検出する。そして、不変化命令検出部620は、部分冗長除去部610から受け取ったプログラムを、検出結果と共に、命令除去部630に送る。

#### 【0034】

命令除去部630は、代数簡約部634と、部分不用代入文除去部638を有する。代数簡約部634は、不変化命令検出部620からプログラムを受け取ると、代数簡約的手法により、可変化命令を除去する。例えば、代数簡約部634は、「可変文字列変数tを不变文字列変数に変換した結果がsであるならば、不变文字列変数sを変換した結果はtになる」という代数的性質を用いて、可変化命令を除去する。具体的には、代数簡約部634は、不変化命令から可変化命令までの間に実行される命令が、不変化命令の変換元の可変文字列変数に格納された文字列を書き換えず、かつ可変化命令から可変化命令により変換された可変文字列変数の使用の間に実行される命令が、不変化命令の変換元の可変文字列変数と、可変化命令により変換された可変文字列変数の何れも書き換えない場合に、可変化命令を除去する。そして、代数簡約部634は、可変化命令より後で、不変化命令の変換元となる可変文字列変数を、可変化命令により変換された可変文字列変数として使用させる。その後、代数簡約部634は、可変化命令を除去したプログラムを部分不用代入文除去部638に送る。

#### 【0035】

部分不用代入文除去部638は、部分不用代入文除去の処理として、以下の処理を行う。部分不用代入文除去部638は、不変化命令を、当該不変化命令より後に実行される分岐命令の分岐先のそれぞれに移動する。そして、部分不用代入文除去部638は、当該分岐命令の分岐先のそれぞれにおいて、不変化命令の変換先である不变文字列変数に格納されている文字列が参照されない場合に、当該不変化命令を除去する。そして、部分不用代入文除去部638は、部分不用代入文除去の処理を行ったプログラムを、逆意味復元部640に送る。

#### 【0036】

逆意味復元部640は、不変化命令検出部620により変換された不変化命令を、不変化命令の機能を実現する複数の命令の組合せに戻す変換処理を行う。そ

して、逆意味復元部640は、変換結果のプログラムを可変文字列変数初期長指定部650に送る。

#### 【0037】

計測コード付きプログラム実行部642は、コンパイラ装置10が行うコンパイルの効率を高めるべく、コンパイル対象のプログラムを予め実行し、実行の状態を計測する。例えば、計測コード付きプログラム実行部642は、文字列変数に実際に格納された文字列の長さを計測し、計測結果を不变文字列変数最終長データベース645に格納する。

#### 【0038】

可変文字列変数初期長指定部650は、コンパイル対象のプログラムを逆意味復元部640から受け取り、可変文字列変数に実際に格納された文字列の長さを示すサイズ情報を不变文字列変数最終長データベース645から受け取る。そして、可変文字列変数初期長指定部650は、可変文字列変数が新たに生成される場合には、サイズ情報が示す長さの領域を予め確保するように、可変文字列変数の初期長を設定する。これにより、コンパイラ装置10は、プログラムの実行時に、可変文字列変数のメモリ領域を伸張させることを防ぐことができる。そして、可変文字列変数初期長指定部650は、可変文字列変数の初期長を設定したプログラムを、コンパイル結果のプログラムとして出力する。

なお、コンパイラ装置60は、計測コード付きプログラム実行部642と、不变文字列変数最終長データベース645と、可変文字列変数初期長指定部650とを備えなくともよい。この場合、コンパイラ装置60は、コンパイラ装置60の設計者により予め設定された値を、可変文字列変数の初期長としてもよい。

#### 【0039】

図7は、コンパイラ装置60の動作フローを示す。意味復元部600は、可変化命令の機能を実現する複数の命令の組合せを検出し、当該命令の組合せを、可変化命令を示す擬似命令に変換する意味復元を行うことにより、不变文字列変数を可変文字列変数に変換する可変化命令を検出する(S700)。部分冗長除去部610は、擬似命令を含むプログラムを意味復元部600から受け取ると、意味復元部600により検出された可変化命令を、当該可変化命令より前において

合流する制御フローの合流元のそれぞれに移動する部分冗長性の除去処理を行う（S710）。不変化命令検出部620は、部分冗長除去部610から受け取ったプログラムにおいて、可変文字列変数を不变文字列変数に変換する不変化命令を検出する（S720）。

#### 【0040】

代数簡約部634は、代数簡約的手法により、可変化命令を除去する。例えば、代数簡約部634は、不変化命令から可変化命令までの間に実行される命令が、不変化命令の変換元の可変文字列変数に格納された文字列を書き換えず、かつ可変化命令から可変化命令により変換された可変文字列変数の使用の間に実行される命令が、不変化命令の変換元の可変文字列変数と、可変化命令により変換された可変文字列変数の何れも書き換えない場合に、可変化命令を除去する。そして、代数簡約部634は、不変化命令の変換元となる可変文字列変数を、除去した可変化命令が存在していた個所より後の部分で、可変化命令により変換された可変文字列変数として使用させる（S730）。

#### 【0041】

部分不用代入文除去部638は、不変化命令を、当該不変化命令より後に実行される分岐命令の分岐先のそれぞれにおいて実行させるように、命令の実行順序を変更する。そして、部分不用代入文除去部638は、当該分岐命令の分岐先のそれぞれにおいて、不変化命令の変換先である不变文字列変数に格納されている文字列が参照されない場合に、当該不変化命令を除去する部分不用代入文除去の処理を行う（S740）。

#### 【0042】

逆意味復元部640は、意味復元部600により変換された可変化命令を、可変化命令の機能を実現する複数の命令の組合せに戻す変換処理を行う（S750）。可変文字列変数初期長指定部650は、計測コード付プログラム実行部642により予め計測され、可変文字列変数に実際に格納された文字列の長さを示すサイズ情報に基づき、可変文字列変数の初期長を設定する（S760）。

#### 【0043】

図8は、コンパイラ装置60が最適化する対象のプログラムの一例を示す。本

図のプログラムは、1から3行目の命令を含む基本ブロック800と、4から9行目の命令を含む基本ブロック810と、10行目の命令を含む基本ブロック820とを有する。

#### 【0044】

1行目の命令は、不变文字列変数の一例であるsに、空の文字列“”を代入する命令である。2行目の命令は、整数型変数のiに、iの初期値を示す変数m inを代入する命令である。3行目の命令は、iと、iの最終値を示すmaxとを比較し、iがmax以上の場合に、基本ブロック820を実行し、iがmaxより小さい場合に、基本ブロック810を実行する分岐命令である。

#### 【0045】

4行目の命令は、可変文字列変数を含むStringBufferオブジェクトtのメモリ上の領域を、新たに確保する命令である。5行目の命令は、sに格納されている文字列をtに追加する命令である。6行目の命令は、文字列型配列変数Aにおける、i番目の文字列を、tに追加する命令である。7行目の命令は、tを、不变文字列変数のsに変換する不変化命令である。8行目の命令は、iをインクリメントする命令である。9行目の命令は、iと、iの最終値を示すmaxとを比較し、iがmax以上の場合に、基本ブロック820を実行し、iがmaxより小さい場合に、再度基本ブロック810を実行する分岐命令である。

#### 【0046】

10行目の命令は、sに格納されている文字列を、標準出力、例えば、ディスプレイ画面のコンソールウィンドウに出力する命令である。

#### 【0047】

図9は、意味復元部600が可変化命令を検出したプログラムの一例を示す。意味復元部600は、図8における4行目及び5行目の命令、即ち、可変文字列変数として用いられる記憶領域を確保する命令と、可変文字列変数に不变文字列変数を複写する命令との組合せを、可変化命令として検出する。そして、意味復元部600は、図9の4行目に示したように、検出した可変化命令を、sからtへの可変化命令である旨を示す擬似命令に置き換える。即ち、意味復元部600

は、可変化命令の処理を構成する複数の命令の組合せを検出し、当該複数の命令が示す処理の意味を復元する。

#### 【0048】

なお、意味復元部600は、プログラムにおいて、連続して配置された4行目及び5行目の命令を検出することにより、可変化命令を検出する。これに代えて、意味復元部600は、4行目及び5行目の命令の間に他の命令が配置されている場合であっても、新たに確保された可変文字列変数に対する最初の操作が不变文字列の追加であれば、4行目及び5行目の命令を、不变化命令として検出してよい。

#### 【0049】

図10は、部分冗長除去部610が部分冗長性を除去したプログラムの一例を示す。部分冗長除去部610は、意味復元部600により検出された可変化命令を、当該可変化命令より前において合流する制御フローの合流元のそれぞれに移動するべく、4行目の可変化命令を、基本ブロック810から除去すると共に、基本ブロック830及び基本ブロック840に挿入する。なお、より詳細には、部分冗長除去部610は、部分冗長性の除去処理として、非特許文献2に示した技術等と略同一の処理を行ってよい。例えば、既存の技術によれば、部分冗長除去部610は、可変化命令を、当該可変化命令による変換元である変数sに値を代入する命令より先に実行させないという条件の元、可変化命令をより先に移動する。

#### 【0050】

図11は、命令除去部630が可変化命令を除去したプログラムの一例を示す。代数簡約部634は、代数簡約的手法により、基本ブロック840に移動した可変化命令を削除する。即ち、基本ブロック810から基本ブロック840に至る制御フローにおいて、図10における6行目の不变化命令及び10行目の可変化命令の間に実行される命令が、可変文字列変数tに格納された文字列を書き換えず、かつ可変化命令から可変化命令により変換された可変文字列変数の使用の間に実行される命令が、不变化命令の変換元の可変文字列変数と、可変化命令により変換された可変文字列変数の何れも書き換えないで、代数簡約部634は

、基本ブロック840における可変化命令を除去する。そして、代数簡約部634は、可変化命令より後で、不変化命令の変換元となる可変文字列変数を、可変化命令により変換された可変文字列変数として使用させる（10行目の命令）。

#### 【0051】

より詳細には、代数簡約部634は、変数に対する値の定義と、変数に格納された値の使用との関係の解析結果、例えば、UD、DUチェインを用いて、可変文字列変数tに格納された文字列が書き換わるか否かを判断する。更に、代数簡約部634は、可変文字列変数tが、他のスレッドからアクセスされないことを更に条件として、可変文字列変数tに格納された文字列が書き換わらないと判断する。一例としては、代数簡約部634は、プログラムが单一のスレッドにより実行されることを条件としてもよいし、可変文字列変数tへのポインタを他のスレッドに通知していないことをエスケープアナリシスにより解析してもよい。

#### 【0052】

図12は、命令除去部630が不変化命令を除去したプログラムの一例を示す。部分不用代入文除去部638は、部分不用代入文の除去処理として、以下の処理を行う。部分不用代入文除去部638は、図11の6行目の不変化命令を、不変化命令より後に実行される分岐命令、例えば7行目の命令の分岐先のそれぞれにおいて実行させる。即ち、部分不用代入文除去部638は、図11の6行目の不変化命令を、基本ブロック810から除去すると共に、基本ブロック840及び基本ブロック850に移動する。そして、部分不用代入文除去部638は、基本ブロック840に移動した不変化命令の変換先である不变文字列変数sに格納されている文字列が、基本ブロック840以降に実行される部分において参照されないので、基本ブロック840に移動した当該不変化命令を除去する。

#### 【0053】

なお、より詳細には、部分不用代入文除去部638は、部分不用代入文の除去処理として、非特許文献1に示した技術等と略同一の処理を行ってもよい。例えば、部分不用代入文除去部638は、不変化命令を、当該不変化命令による変換先である変数sを参照する命令より後に実行させないという条件の元、不変化命令をより後に移動する。

## 【0054】

図13は、コンパイラ装置10又はコンパイラ装置60のハードウェア構成の一例を示す。第1の実施形態又は変形例に係るコンパイラ装置10並びに第2の実施形態に係るコンパイラ装置60は、ホストコントローラ1082により相互に接続されるCPU1000、RAM1020、グラフィックコントローラ1075、及び表示装置1080を有するCPU周辺部と、入出力コントローラ1084によりホストコントローラ1082に接続される通信インターフェイス1030、ハードディスクドライブ1040、及びCD-ROMドライブ1060を有する入出力部と、入出力コントローラ1084に接続されるROM1010、フレキシブルディスクドライブ1050、及び入出力チップ1070を有するレガシー入出力部とを備える。

## 【0055】

ホストコントローラ1082は、RAM1020と、高い転送レートでRAM1020をアクセスするCPU1000及びグラフィックコントローラ1075とを接続する。CPU1000は、ROM1010及びRAM1020に格納されたコンパイラプログラムに基づいて動作し、各部の制御を行う。グラフィックコントローラ1075は、CPU1000等がRAM1020内に設けたフレームバッファ上に生成する画像データを取得し、表示装置1080上に表示させる。これに代えて、グラフィックコントローラ1075は、CPU1000等が生成する画像データを格納するフレームバッファを、内部に含んでもよい。

## 【0056】

入出力コントローラ1084は、ホストコントローラ1082と、比較的高速な入出力装置である通信インターフェイス1030、ハードディスクドライブ1040、及びCD-ROMドライブ1060を接続する。通信インターフェイス1030は、ネットワークを介して他の装置と通信する。ハードディスクドライブ1040は、コンパイラ装置10又はコンパイラ装置60が使用するコンパイラプログラム及びデータを格納する。CD-ROMドライブ1060は、CD-ROM1095からコンパイラプログラム又はデータを読み取り、RAM1020を介して入出力チップ1070に提供する。

## 【0057】

また、入出力コントローラ1084には、ROM1010と、フレキシブルディスクドライブ1050や入出力チップ1070等の比較的低速な入出力装置とが接続される。ROM1010は、コンパイラ装置10又はコンパイラ装置60の起動時にCPU1000が実行するブートプログラムや、コンパイラ装置10又はコンパイラ装置60のハードウェアに依存するプログラム等を格納する。フレキシブルディスクドライブ1050は、フレキシブルディスク1090からコンパイラプログラム又はデータを読み取り、RAM1020を介して入出力チップ1070に提供する。入出力チップ1070は、フレキシブルディスク1090や、例えばパラレルポート、シリアルポート、キーボードポート、マウスポート等を介して各種の入出力装置を接続する。

## 【0058】

コンパイラ装置10又はコンパイラ装置60に提供されるコンパイラプログラムは、フレキシブルディスク1090、CD-ROM1095、又はICカード等の記録媒体に格納されて利用者によって提供される。コンパイラプログラムは、記録媒体から読み出され、入出力チップ1070を介してコンパイラ装置10又はコンパイラ装置60にインストールされ、コンパイラ装置10又はコンパイラ装置60において実行される。

## 【0059】

コンパイラ装置10又はコンパイラ装置60にインストールされて実行されるコンパイラプログラムは、追加命令検出モジュールと、格納コード生成モジュールと、参照命令検出モジュールと、追加コード生成モジュールと、不変化命令検出モジュールと、部分冗長除去モジュールと、可変化命令検出モジュールと、命令除去モジュールと、代数簡約モジュールと、部分不用代入文除去モジュールと、逆意味復元モジュールと、計測コード付きプログラム実行モジュールと、可変文字列変数初期長指定モジュールとを含む。各モジュールがコンパイラ装置10又はコンパイラ装置60に働きかけて行わせる動作は、図1から図12において説明したコンパイラ装置10又はコンパイラ装置60における、対応する部材の動作と同一であるから、説明を省略する。

## 【0060】

以上に示したコンパイラプログラム又はモジュールは、外部の記憶媒体に格納されてもよい。記憶媒体としては、フレキシブルディスク1090、CD-R0M1095の他に、DVDやPD等の光学記録媒体、MD等の光磁気記録媒体、テープ媒体、ICカード等の半導体メモリ等を用いることができる。また、専用通信ネットワークやインターネットに接続されたサーバシステムに設けたハードディスク又はRAM等の記憶装置を記録媒体として使用し、ネットワークを介してコンパイラプログラムをコンパイラ装置10又はコンパイラ装置60に提供してもよい。

## 【0061】

以上の説明から明らかなように、コンパイラ装置は、文字列を追加する命令を最適化することができる。

## 【0062】

以上、本発明を実施形態を用いて説明したが、本発明の技術的範囲は上記実施形態に記載の範囲には限定されない。上記実施形態に、多様な変更または改良を加えることができる。そのような変更または改良を加えた形態も本発明の技術的範囲に含まれ得ることが、特許請求の範囲の記載から明らかである。例えば、コンパイラ装置10は、コンパイラ装置60の各部材を更に備えてよい。この場合、コンパイラ装置10は、予め定められた条件に基づき、第1実施形態及び第2実施形態に示した機能の何れかを選択して処理してもよい。例えば、コンパイラ装置10は、計測コード付きプログラム実行部642により文字列の長さが計測できた場合は、第2実施形態に示した処理を行い、計測できなかった場合は、第1実施形態に示した処理を行ってよい。また、コンパイラ装置10は、計測コード付きプログラム実行部642により計測された複数の文字列の長さの分散値が、所定値より大きい場合に、第1実施形態に示した処理を行ってよい。

## 【0063】

以上で示した全ての実施形態及び変形例によると、以下の各項目に示すコンパイラ装置、コンパイラプログラム、及び記録媒体が実現される。

## 【0064】

(項目1) 文字列を操作するプログラムを最適化するコンパイラ装置であって、前記プログラムにおいて、文字列を格納する文字列変数に文字列を追加する追加命令を検出する追加命令検出部と、前記追加命令検出部により検出された、同一の文字列変数に文字列を追加する複数の追加命令のそれぞれに代えて、当該追加命令により追加される追加文字列のデータをバッファに格納する格納コードを生成する格納コード生成部と、前記プログラムにおける、前記文字列変数を参照する命令より先に実行される個所に、複数の前記追加文字列のそれぞれを前記文字列変数に追加する追加コードを生成する追加コード生成部とを備えるコンパイラ装置。

(項目2) 前記複数の追加命令によって文字列を追加した後に初めて前記文字列変数を参照する参照命令を検出する参照命令検出部を更に備え、前記追加コード生成部は、前記追加コードを、全ての前記格納コードより後、かつ前記参照命令より先に実行される個所に生成する項目1記載のコンパイラ装置。

#### 【0065】

(項目3) 前記追加命令検出部は、前記追加命令として、文字列を追加する処理が許可されていない不变文字列変数を、文字列を追加する処理が許可された可変文字列変数に変換する命令と、前記可変文字列変数に前記追加文字列を追加する命令と、前記可変文字列変数を不变文字列変数に変換する命令との組合せを検出する項目1記載のコンパイラ装置。

(項目4) 文字列を操作するプログラムを最適化するコンパイラ装置であって、前記プログラムにおいて、文字列を格納する文字列変数に文字列を追加する追加命令を検出する追加命令検出部と、前記追加命令検出部により検出された、同一の文字列変数に文字列を追加する複数の追加命令のそれぞれに代えて、当該追加命令により追加される追加文字列が格納されているメモリのアドレスをバッファに格納する格納コードを生成する格納コード生成部と、前記プログラムにおける、前記文字列変数を参照する命令より先に実行される個所に、複数の前記アドレスに格納される複数の追加文字列のそれぞれを前記文字列変数に追加する追加コードを生成する追加コード生成部とを備えるコンパイラ装置。

#### 【0066】

(項目5) 文字列を操作するプログラムを最適化するコンパイラ装置であって、文字列を追加する処理が許可された可変文字列変数を、文字列を追加する処理が許可されていない不变文字列変数に変換する不変化命令を検出する不変化命令検出部と、前記不变文字列変数を可変文字列変数に変換する可変化命令を検出する可変化命令検出部と、前記不変化命令から前記可変化命令までの間に実行される命令が、前記不変化命令の変換元の可変文字列変数に格納された文字列を書き換えず、かつ前記可変化命令から前記可変化命令により変換された可変文字列変数の使用の間に実行される命令が、前記不変化命令の変換元の可変文字列変数と、前記可変化命令により変換された可変文字列変数の何れも書き換えない場合に、前記可変化命令を除去し、前記可変化命令より後で、前記不変化命令の変換元となる可変文字列変数を、前記可変化命令により変換された可変文字列変数として使用させる命令除去部とを備えるコンパイラ装置。

(項目6) 前記不变文字列変数に格納された文字列が参照されない場合に、前記命令除去部は、前記不変化命令を更に除去する項目5記載のコンパイラ装置。

(項目7) 前記命令除去部は、前記不変化命令を、当該不変化命令より後に実行される分岐命令の分岐先のそれぞれに移動し、当該分岐命令の分岐先のそれぞれにおいて、不変化命令の変換先である不变文字列変数に格納されている文字列が参照されない場合に、当該不変化命令を除去する部分不用代入文除去を行う項目6記載のコンパイラ装置。

#### 【0067】

(項目8) 前記可変化命令検出部は、可変文字列変数として用いられる記憶領域を確保する命令と、当該可変文字列変数に、前記不变文字列変数に格納された文字列を追加する命令との組合せを、前記可変化命令として検出する項目5記載のコンパイラ装置。

(項目9) 前記可変化命令検出部により検出された前記可変化命令を、当該可変化命令より前において合流する制御フローの合流元のそれぞれに移動する部分冗長性の除去処理を行う部分冗長除去部を更に備え、前記部分冗長性の前記除去処理が実行されたプログラムにおいて、前記不変化命令及び前記可変化命令の間に実行される命令が、前記不変化命令の変換元の可変文字列変数に格納された文

字列を書き換えず、かつ前記可変化命令から前記可変化命令により変換された可変文字列変数の使用の間に実行される命令が、前記不変化命令の変換元の可変文字列変数と、前記可変化命令により変換された可変文字列変数の何れも書き換えない場合に、前記命令除去部は、前記可変化命令を除去する項目5記載のコンパイラ装置。

## 【0068】

(項目10) 前記命令除去部は、前記不変化命令を、当該不変化命令より後に実行される分岐命令の分岐先のそれぞれに移動し、当該分岐命令の分岐先のそれぞれにおいて、不変化命令の変換先である不変文字列変数に格納されている文字列が参照されない場合に、当該不変化命令を除去する部分不用代入文除去を行う項目9記載のコンパイラ装置。

(項目11) 文字列を操作するプログラムをコンピュータにより最適化するコンパイラプログラムであって、前記コンピュータを、前記プログラムにおいて、文字列を格納する文字列変数に文字列を追加する追加命令を検出する追加命令検出部と、前記追加命令検出部により検出された、同一の文字列変数に文字列を追加する複数の追加命令のそれぞれに代えて、当該追加命令により追加される追加文字列のデータをバッファに格納する格納コードを生成する格納コード生成部と、前記プログラムにおける、前記文字列変数を参照する命令より先に実行される個所に、複数の前記追加文字列のそれぞれを前記文字列変数に追加する追加コードを生成する追加コード生成部として機能させるコンパイラプログラム。

## 【0069】

(項目12) 文字列を操作するプログラムをコンピュータにより最適化するコンパイラプログラムであって、前記コンピュータを、前記プログラムにおいて、文字列を格納する文字列変数に文字列を追加する追加命令を検出する追加命令検出部と、前記追加命令検出部により検出された、同一の文字列変数に文字列を追加する複数の追加命令のそれぞれに代えて、当該追加命令により追加される追加文字列が格納されているメモリのアドレスをバッファに格納する格納コードを生成する格納コード生成部と、前記プログラムにおける、前記文字列変数を参照する命令より先に実行される個所に、複数の前記アドレスに格納される複数の追加

文字列のそれぞれを前記文字列変数に追加する追加コードを生成する追加コード生成部として機能させるコンパイラプログラム。

(項目13) 文字列を操作するプログラムをコンピュータにより最適化するコンパイラプログラムであって、前記コンピュータを、文字列を追加する処理が許可された可変文字列変数を、文字列を追加する処理が許可されていない不变文字列変数に変換する不変化命令を検出する不変化命令検出部と、前記不变文字列変数を可変文字列変数に変換する可変化命令を検出する可変化命令検出部と、前記不変化命令から前記可変化命令までの間に実行される命令が、前記不変化命令の変換元の可変文字列変数に格納された文字列を書き換えず、かつ前記可変化命令から前記可変化命令により変換された可変文字列変数の使用の間に実行される命令が、前記不変化命令の変換元の可変文字列変数と、前記可変化命令により変換された可変文字列変数の何れも書き換えない場合に、前記可変化命令を除去し、前記可変化命令より後で、前記不変化命令の変換元となる可変文字列変数を、前記可変化命令により変換された可変文字列変数として使用させる命令除去部として機能させるコンパイラプログラム。

(項目14) 項目11から13の何れかに記載のコンパイラプログラムを記録した記録媒体。

#### 【0070】

##### 【発明の効果】

上記説明から明らかなように、本発明によれば文字列を操作するプログラムを最適化することができる。

##### 【図面の簡単な説明】

###### 【図1】

図1は、コンパイラ装置10のブロック図を示す。

###### 【図2】

図2は、コンパイラ装置10の動作フローを示す。

###### 【図3】

図3(a)は、最適化対象のプログラムのソースコードの一例を示す。

図3(b)は、プログラムがコンパイルされた結果を示す。

図3 (c) は、コンパイラ装置10による最適化を終えたプログラムの一例を示す。

【図4】

図4は、変形例において、コンパイラ装置10がプログラムを最適化する一例を示す。

【図5】

図5 (a) は、第1の他の方法においてプログラムが最適化された結果を示す。

図5 (b) は、第2の他の方法においてプログラムが最適化された結果を示す。

【図6】

図6は、コンパイラ装置60のブロック図を示す。

【図7】

図7は、コンパイラ装置60の動作フローを示す。

【図8】

図8は、コンパイラ装置60が最適化する対象のプログラムの一例を示す。

【図9】

図9は、意味復元部600が可変化命令を検出したプログラムの一例を示す。

【図10】

図10は、部分冗長除去部610が部分冗長性を除去したプログラムの一例を示す。

【図11】

図11は、命令除去部630が可変化命令を除去したプログラムの一例を示す。

【図12】

図12は、命令除去部630が不変化命令を除去したプログラムの一例を示す。

【図13】

図13は、コンパイラ装置10又はコンパイラ装置60のハードウェア構成の

一例を示す。

【符号の説明】

100 コンパイラ装置

600 コンパイラ装置

100 追加命令検出部

110 格納コード生成部

120 参照命令検出部

130 追加コード生成部

600 可変命令検出部

610 意味復元部

620 不変化命令検出部

630 命令除去部

634 代数簡約部

638 部分不用代入文除去部

640 逆意味復元部

642 計測コード付きプログラム実行部

645 不変文字列変数最終長データベース

650 可変文字列変数初期長指定部

800 基本ブロック

810 基本ブロック

820 基本ブロック

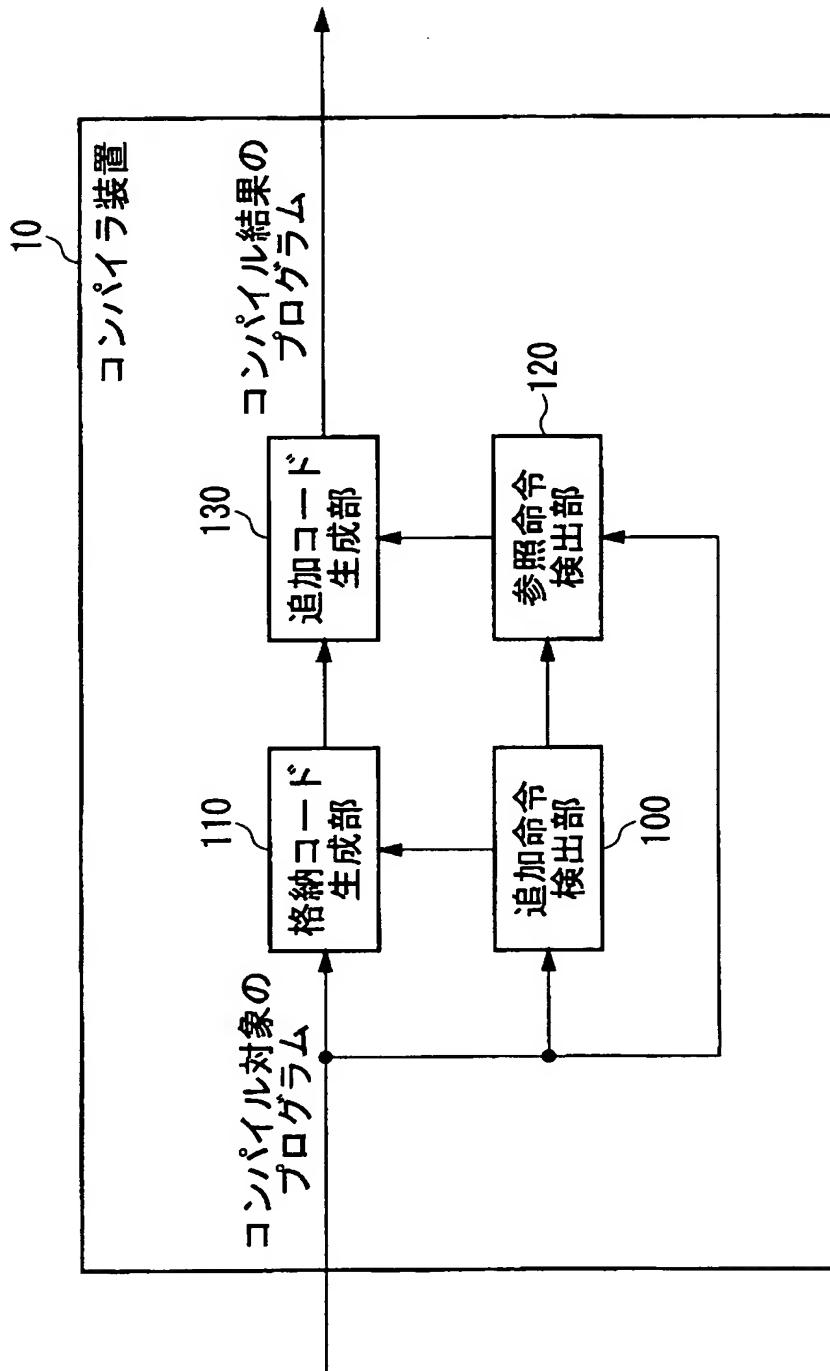
830 基本ブロック

840 基本ブロック

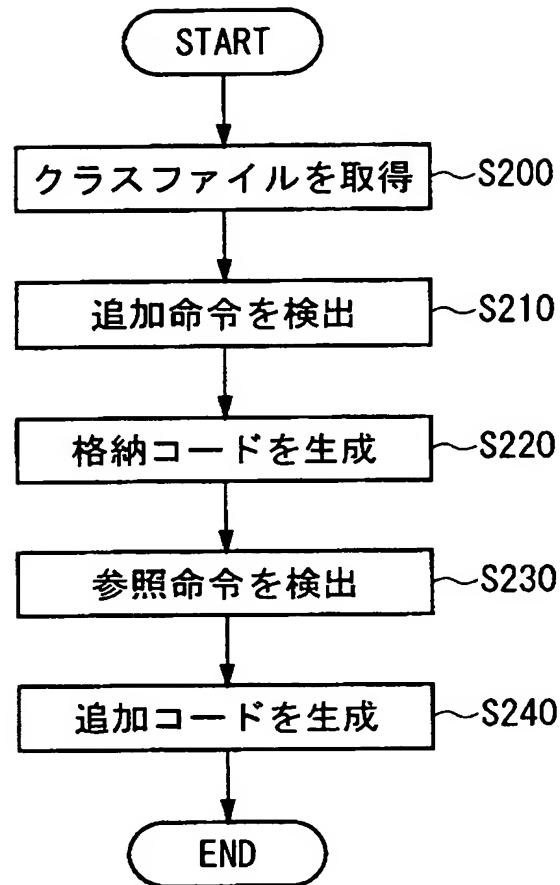
850 基本ブロック

【書類名】 図面

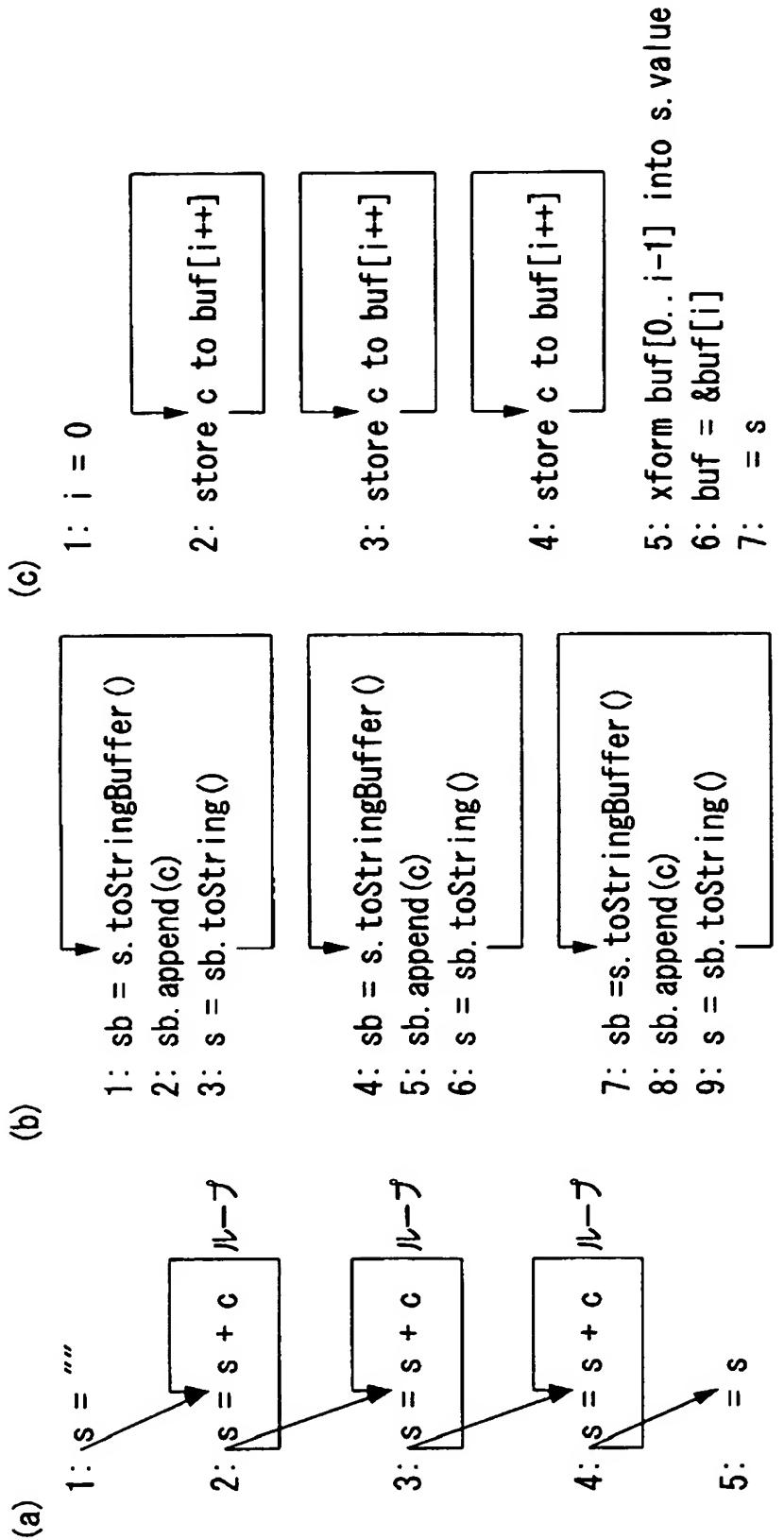
### 【図1】



【図2】



【図3】



【図4】

1: bufの初期化

2: store c to buf[ ]  
3: buf = buf → next4: store c to buf[ ]  
5: buf = buf → next6: store c to buf[ ]  
7: buf = buf → next8: bufのリストをs.valueに変換  
9:     = s

【図5】

(a)

1: sb = s.toStringBuffer()

2: sb.append(c)

3: sb.append(c)

4: sb.append(c)

5: s = sb.toString()

6:     = s

(b)

1: s =

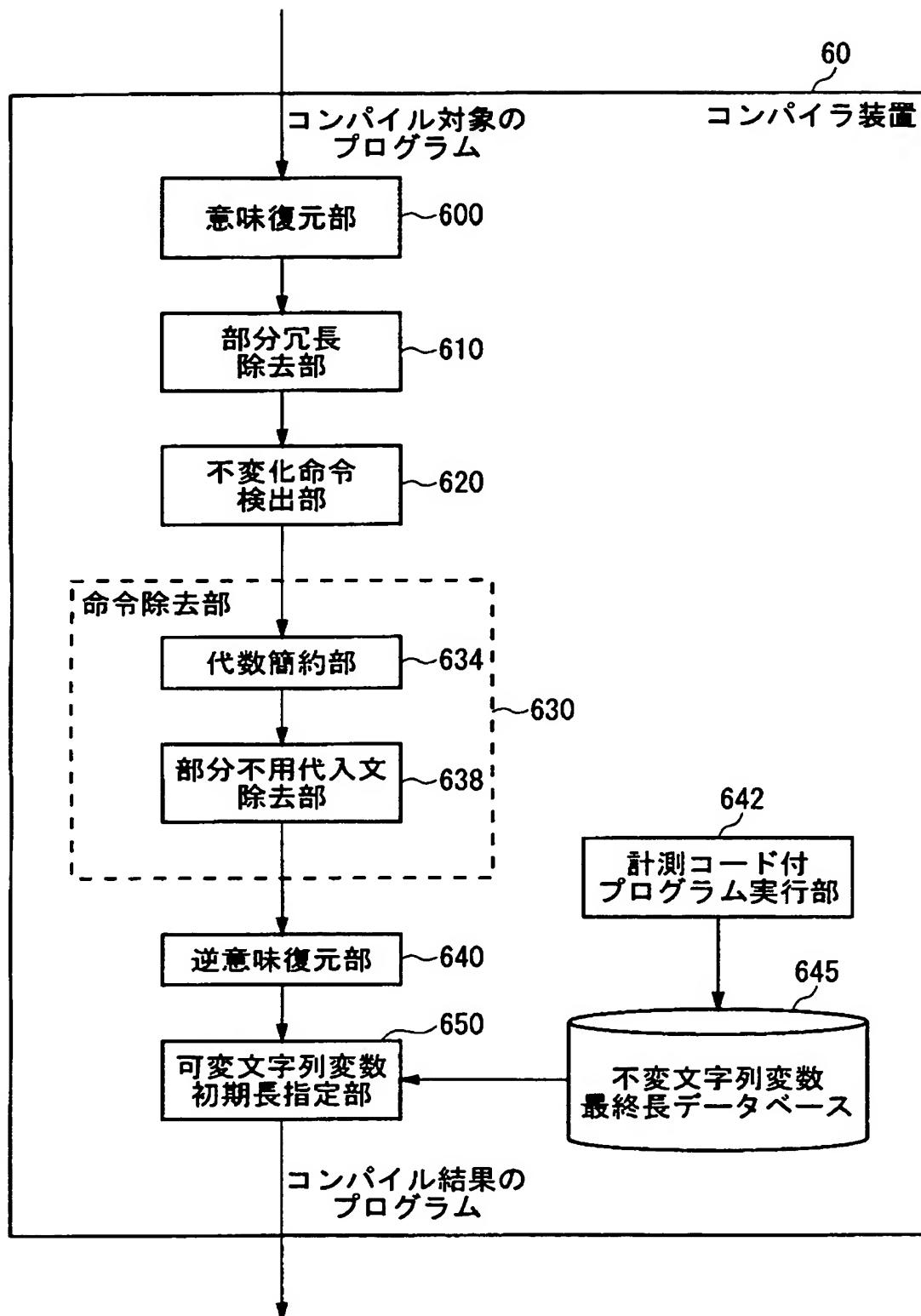
2: s = new String(s, c)

3: s = new String(s, c)

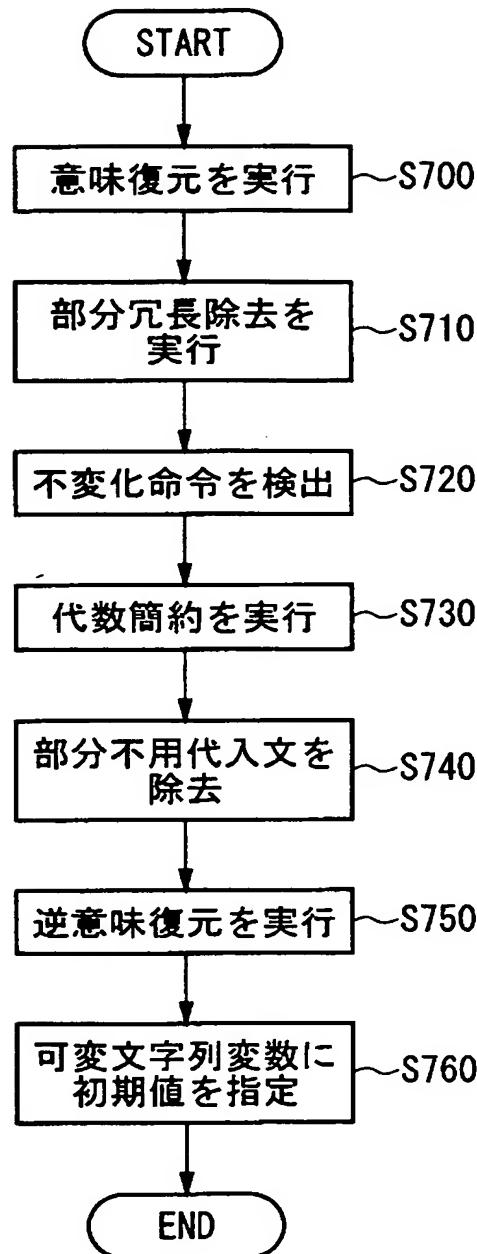
4: s = new String(s, c)

5:     = s

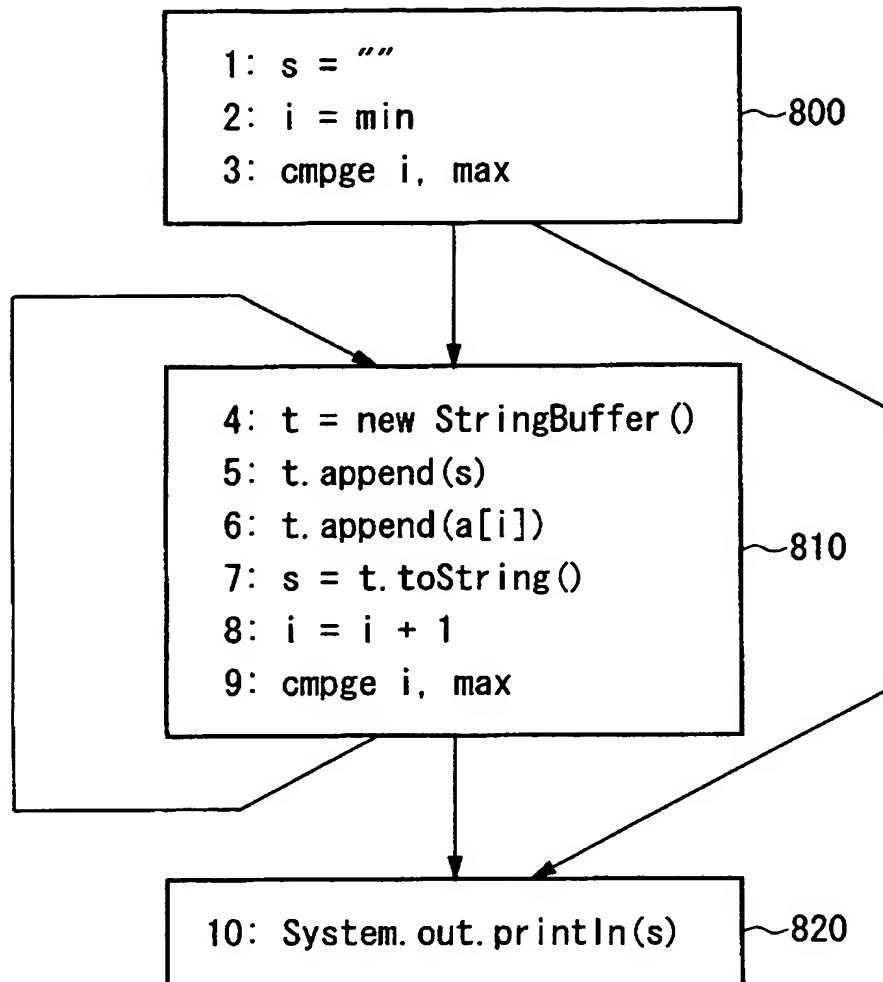
【図6】



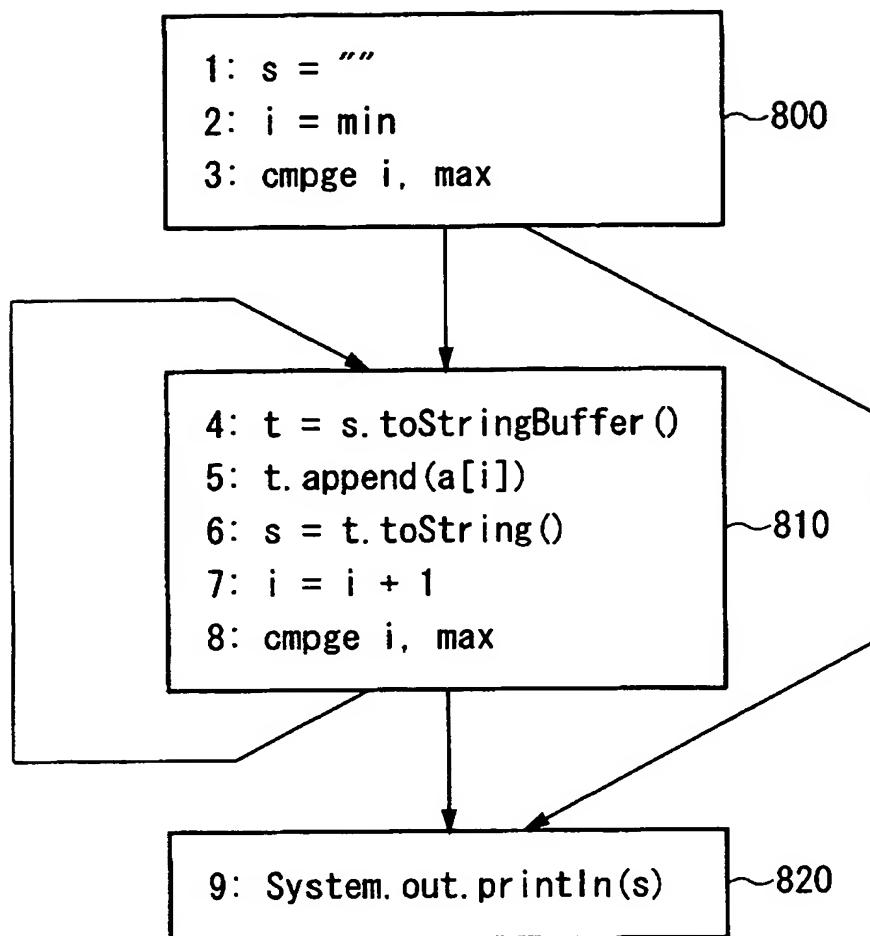
【図7】



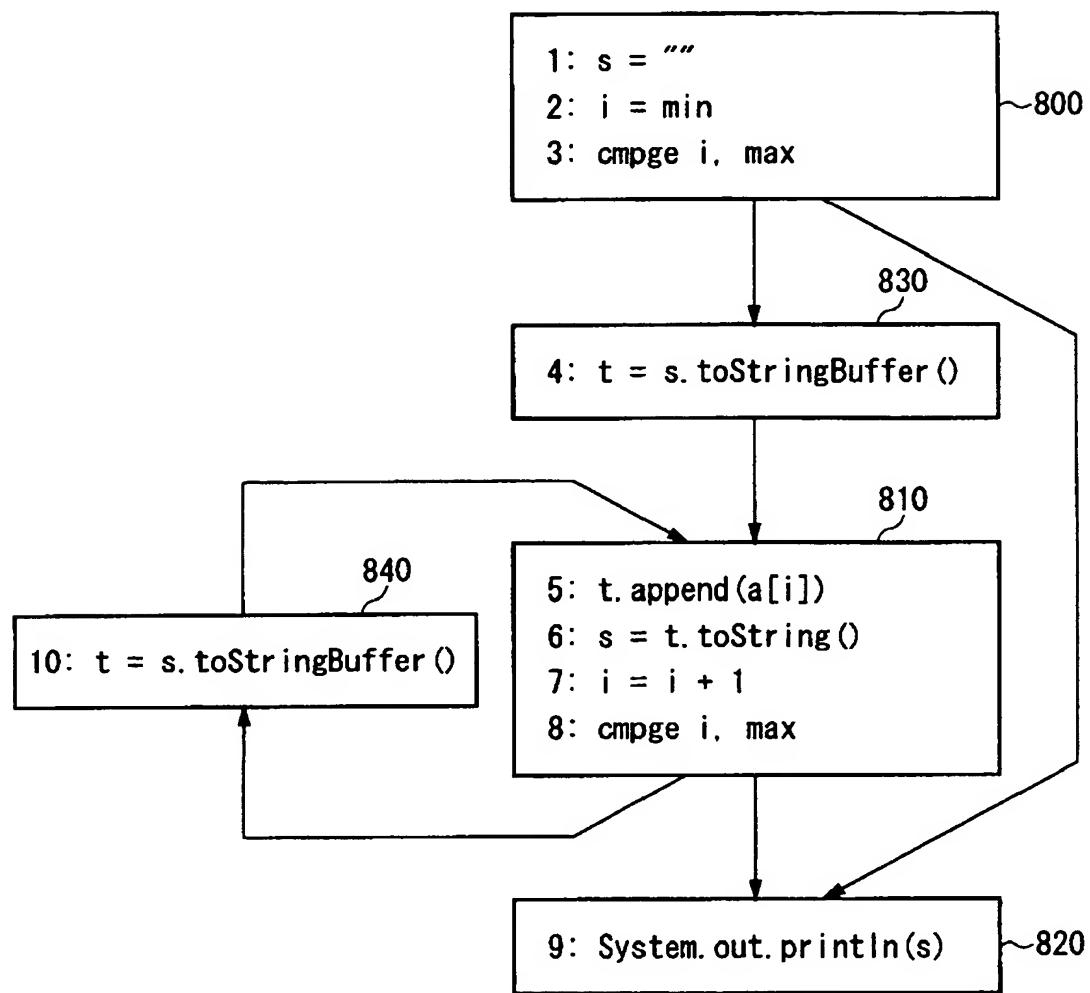
【図8】



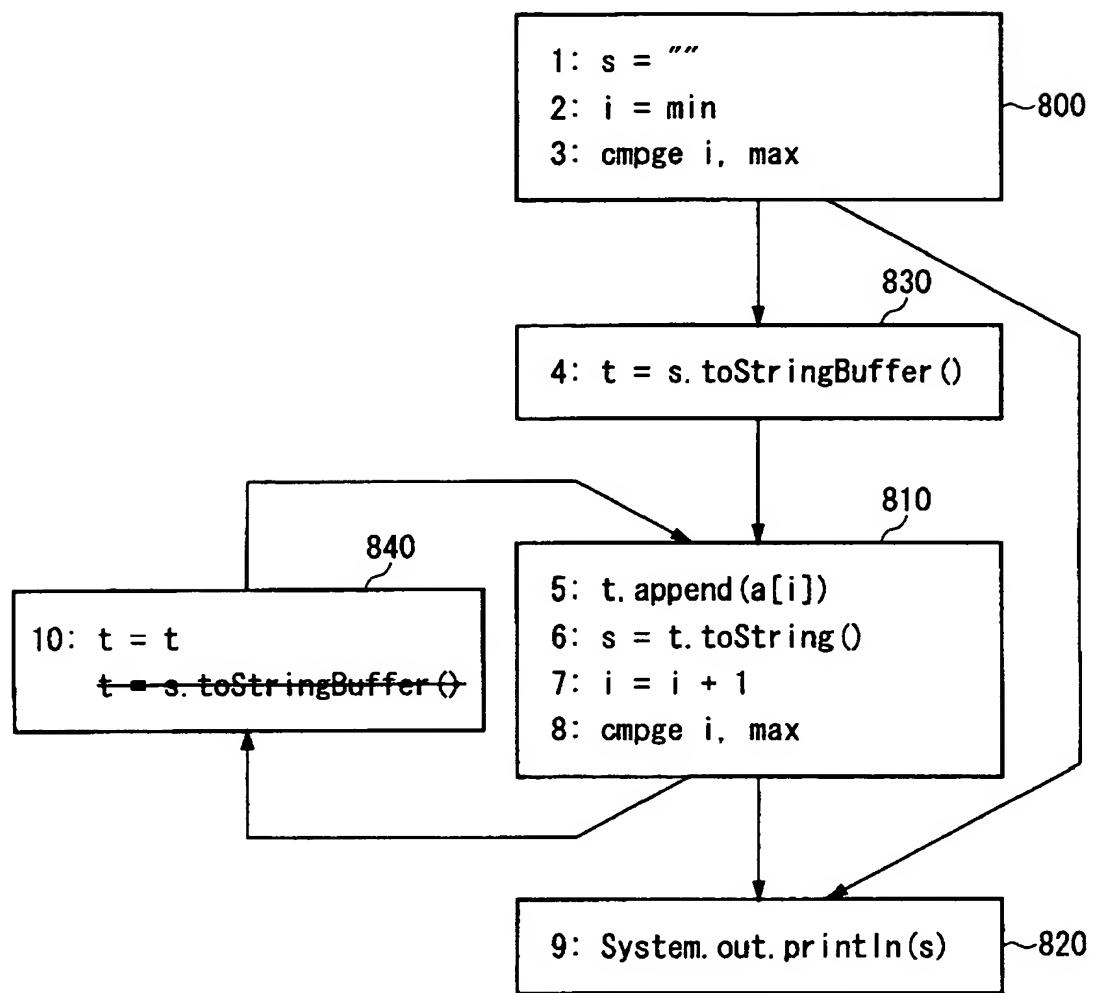
【図9】



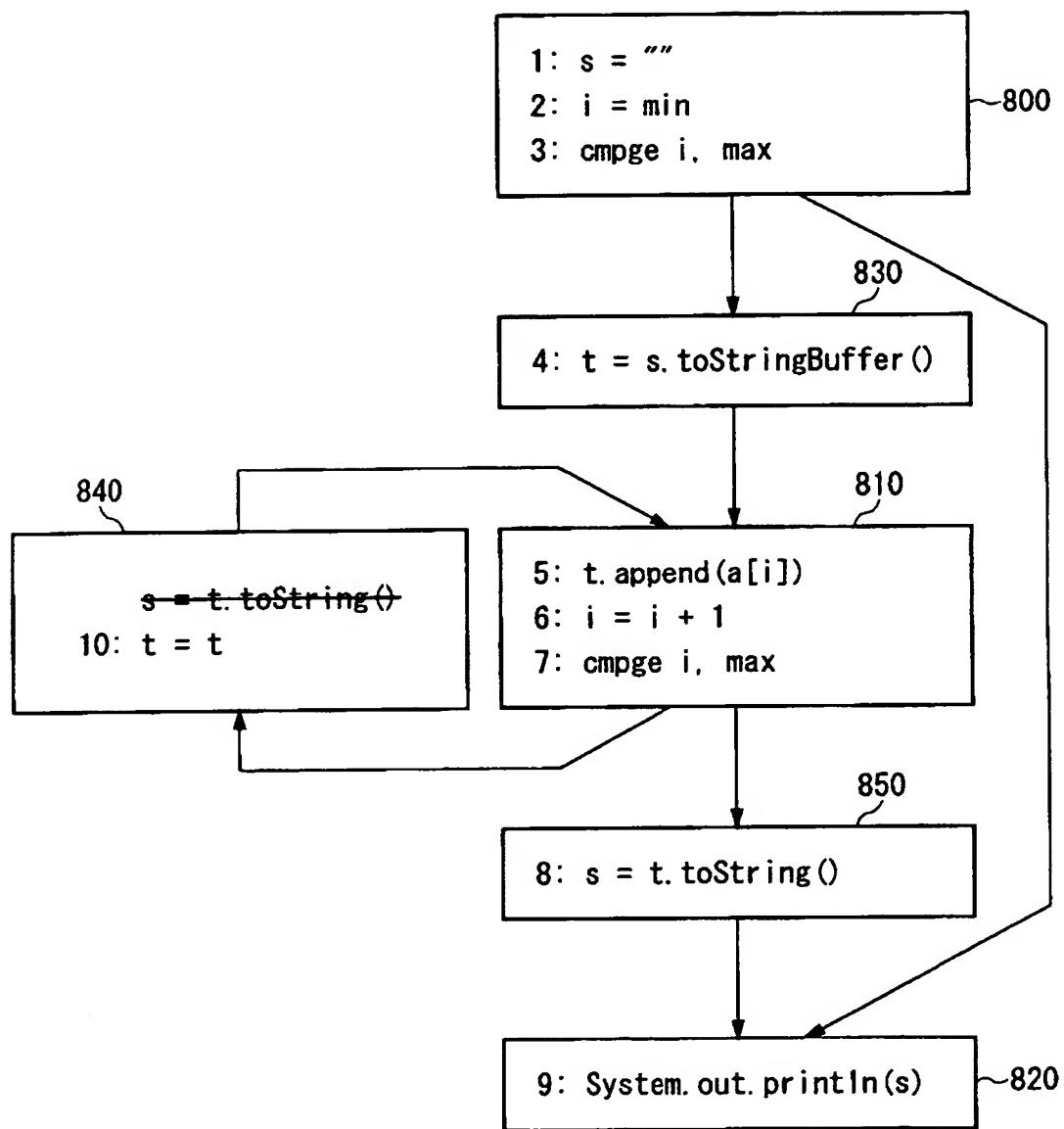
【図10】



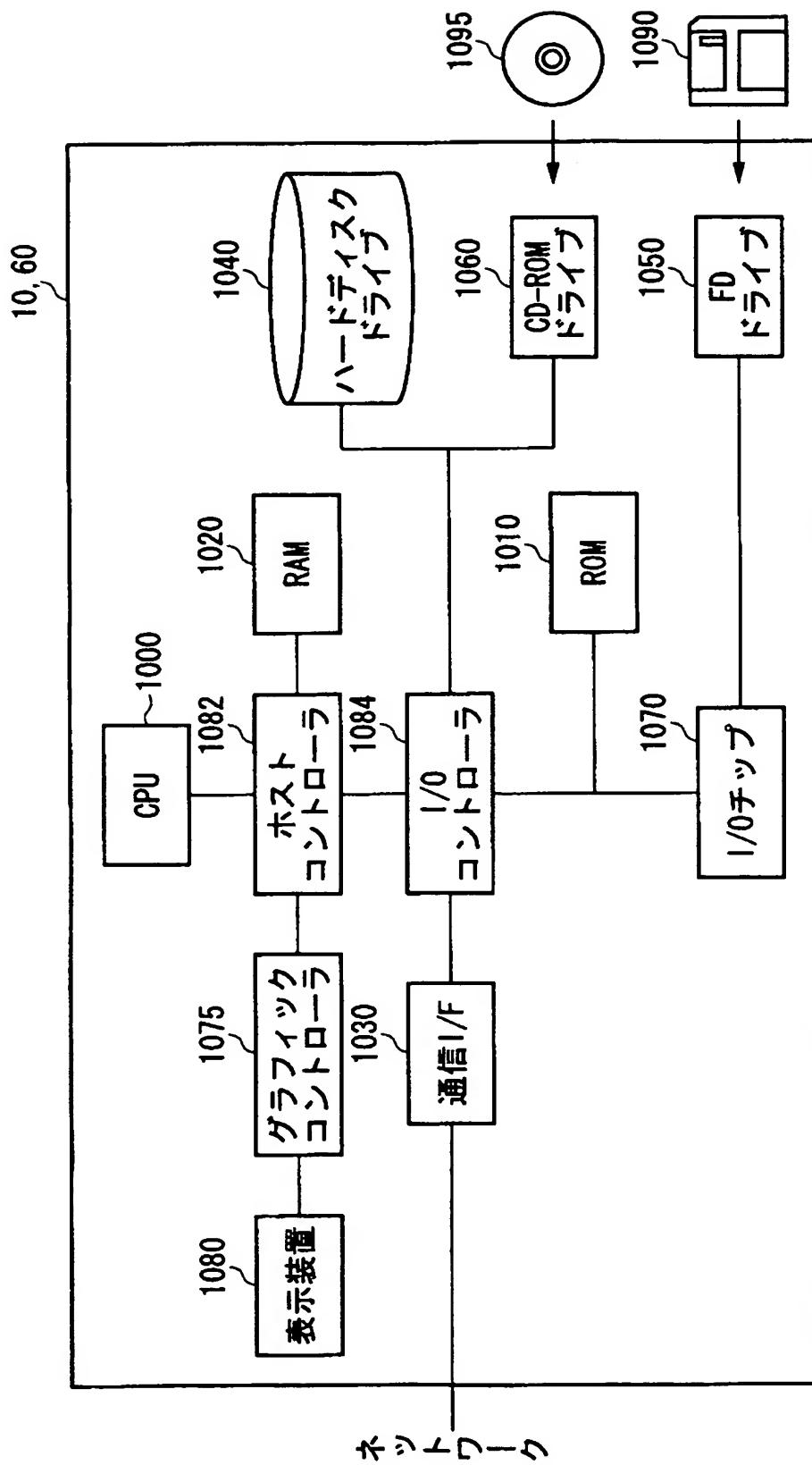
【図11】



【図12】



【図13】



【書類名】 要約書

【要約】

【課題】 文字列を操作するプログラムを最適化する。

【解決手段】 文字列を操作するプログラムを最適化するコンパイラ装置は、プログラムにおいて、文字列を格納する文字列変数に文字列を追加する追加命令を検出する追加命令検出部と、追加命令検出部により検出された、同一の文字列変数に文字列を追加する複数の追加命令のそれぞれに代えて、当該追加命令により追加される追加文字列のデータをバッファに格納する格納コードを生成する格納コード生成部と、プログラムにおける、文字列変数を参照する命令より先に実行される個所に、複数の追加文字列のそれぞれを文字列変数に追加する追加コードを生成する追加コード生成部とを備える。

【選択図】 図1

## 認定・付加情報

特許出願の番号	特願2003-049414
受付番号	50300310530
書類名	特許願
担当官	伊藤 雅美 2132
作成日	平成15年 4月 8日

## &lt;認定情報・付加情報&gt;

## 【特許出願人】

【識別番号】	390009531
【住所又は居所】	アメリカ合衆国 10504、ニューヨーク州 アーモンク ニュー オーチャード ロード
【氏名又は名称】	インターナショナル・ビジネス・マシンズ・コーポレーション

## 【代理人】

【識別番号】	100086243
【住所又は居所】	神奈川県大和市下鶴間 1623 番地 14 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	坂口 博

## 【代理人】

【識別番号】	100091568
【住所又は居所】	神奈川県大和市下鶴間 1623 番地 14 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	市位 嘉宏

## 【代理人】

【識別番号】	100108501
【住所又は居所】	神奈川県大和市下鶴間 1623 番 14 日本アイ・ビー・エム株式会社 知的所有権
【氏名又は名称】	上野 剛史

## 【復代理人】

【識別番号】	100104156
【住所又は居所】	東京都新宿区新宿 1 丁目 24 番 12 号 東信ビル 6 階 龍華国際特許事務所
【氏名又は名称】	龍華 明裕

次頁無

出願人履歴情報

識別番号 [390009531]

1. 変更年月日 2002年 6月 3日

[変更理由] 住所変更

住 所 アメリカ合衆国10504、ニューヨーク州 アーモンク ニュー オーチャード ロード

氏 名 インターナショナル・ビジネス・マシーンズ・コーポレーション